DOCUMENT RESUME

ED 235 233 TM 830 659

AUTHOR Holmes, Susan E.
TITLE CSAR: An Interactive Item Bank System for the Storage and Retrieval of Item Information.
PUB DATE Apr 83
NOTE 46p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Montreal, Quebec, April 12-14, 1983).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Descriptive (141)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Achievement Tests; *Computer Assisted Testing; Information Retrieval; Information Storage; Item Analysis; *Item Banks; Latent Trait Theory; *Online Systems; Test Construction; Test Items
IDENTIFIERS *CSAR Interactive Item Bank System

ABSTRACT

CSAR (Characteristic Storage And Retrieval) is the name of an interactive item bank system for the storage, retrieval, manipulation, and summarization of information about multiple choice test items. The purpose of the system is to facilitate the test construction process. The system has two main components: one or more item banks and an interactive item selection program. Nonstatistical, traditional item analysis, and item response theory item information is stored in each bank. This information is then accessed interactively by test editors using a wide variety of different item selection criteria. The paper describes the CSAR system briefly and discusses the design decisions around which it was developed. Recommendations are offered to others who might be considering a similar venture. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED235233

CSAR: An Interactive Item Bank System for the Storage and Retrieval of Item Information

Susan E. Holmes
Occupational Assessment Group
CTB/McGraw-Hill

A paper presented at a symposium entitled
Item Banking: Practice and Potential in the 1980's
held during the National Council on Measurement in Education Meeting
Montreal, Canada, April 1983

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

S. E. Holmes

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 830 659

Abstract

CSAR is the name of an interactive item bank system for the storage, retrieval, manipulation, and summarization of information about multiple choice test items. The purpose of the system is to facilitate the test construction process. The system has two main components: one or more item banks and an interactive item selection program. Nonstatistical, traditional item analysis, and item response theory item information is stored in each bank. This information is then accessed interactively by test editors using a wide variety of different item selection criteria. The paper describes the CSAR system briefly and discusses the design decisions around which it was developed. Recommendations are offered to others who might be considering a similar venture.

CSAR: An Interactive Item Bank System for the Storage and Retrieval of Item Information[1]

Introduction

CSAR is the name of an interactive item bank system developed by CTB/McGraw-Hill. The system was designed to facilitate the storage, retrieval, manipulation, and summarization of information about multiple choice test items used in the test construction process. The purpose of this paper is to describe the CSAR system, to discuss the design decisions around which it was developed, and to make recommendations to others who might be considering a similar venture.

The acronym CSAR stands for Characteristic Storage And Retrieval. CSAR was originally developed for a specific application in the area of licensing health professionals. CTB/McGraw-Hill currently produces 12 examinations each year for four nurse licensing programs: two registered nurse, two practical nurse, four vocational nurse, and four psychiatric technician examinations. Each examination contains between 250 and 480 multiple choice test items used to determine licensure and a subset of experimental items for use in subsequent examinations. In all, over 180,000 nursing candidates are tested each year and over 5000 test items analyzed. The CSAR system was developed to meet the needs of these programs.

While most licensing examination programs differ with respect to scope and detail, they share many common characteristics. For example, such programs usually involve multiple administrations of equivalent examination forms with equivalent passing scores. The examinations are constructed according to a

---

[1]The author wishes to acknowledge Ron Weitzman and Abacus Information Management Systems for their respective contributions to the design of the CSAR system.

well-developed test plan, including a test content outline to which test items are coded, and experimental items are usually field tested prior to use in an examination. The CSAR system was developed around these common elements.

An overview of the CSAR system is presented in Figure 1. The cycle displayed there represents a single test administration. It begins in the upper left hand corner where an examination that includes some experimental items is administered. Candidate responses to the items in the examination are then scored, and various individual and summary reports are produced. The data from the administration are then analyzed in a variety of ways and stored on disk in an "item bank." The information from this and other administrations is then accessed interactively by test editors who review it and select items for inclusion in some future examination. The new examination is then administered, and the cycle begins again. At any given time throughout the calendar year, one or more of the 12 examinations produced by CTB may be at each step in the cycle.

The CSAR system has two main components: one or more item banks and an interactive item selection program. The first component is an item bank. An item bank is usually defined to be a collection of test items and their characteristics stored together for test construction purposes. In this instance, however, only item characteristics are included in an item bank. The text of test items is stored separately in a text editing system. Although there is no direct link between the two systems currently, the potential for such an interface exists since the same item identification scheme is used in both systems. The item bank is a Virtual Storage Access Method (VSAM) disk file containing one data record for each administration of an item. It is created and updated in batch mode by several subroutines written in the PL/1 programming language.

The second component of the CSAR system is an item selection program which executes under CICS/VS. The program consists of 12 display screens for a 3270 type terminal. It allows one or more test editors to select items from an existing item bank and to then display and summarize information about the items selected. The CSAR system was developed and implemented on the IBM 3031, operating under DOS/VSE.

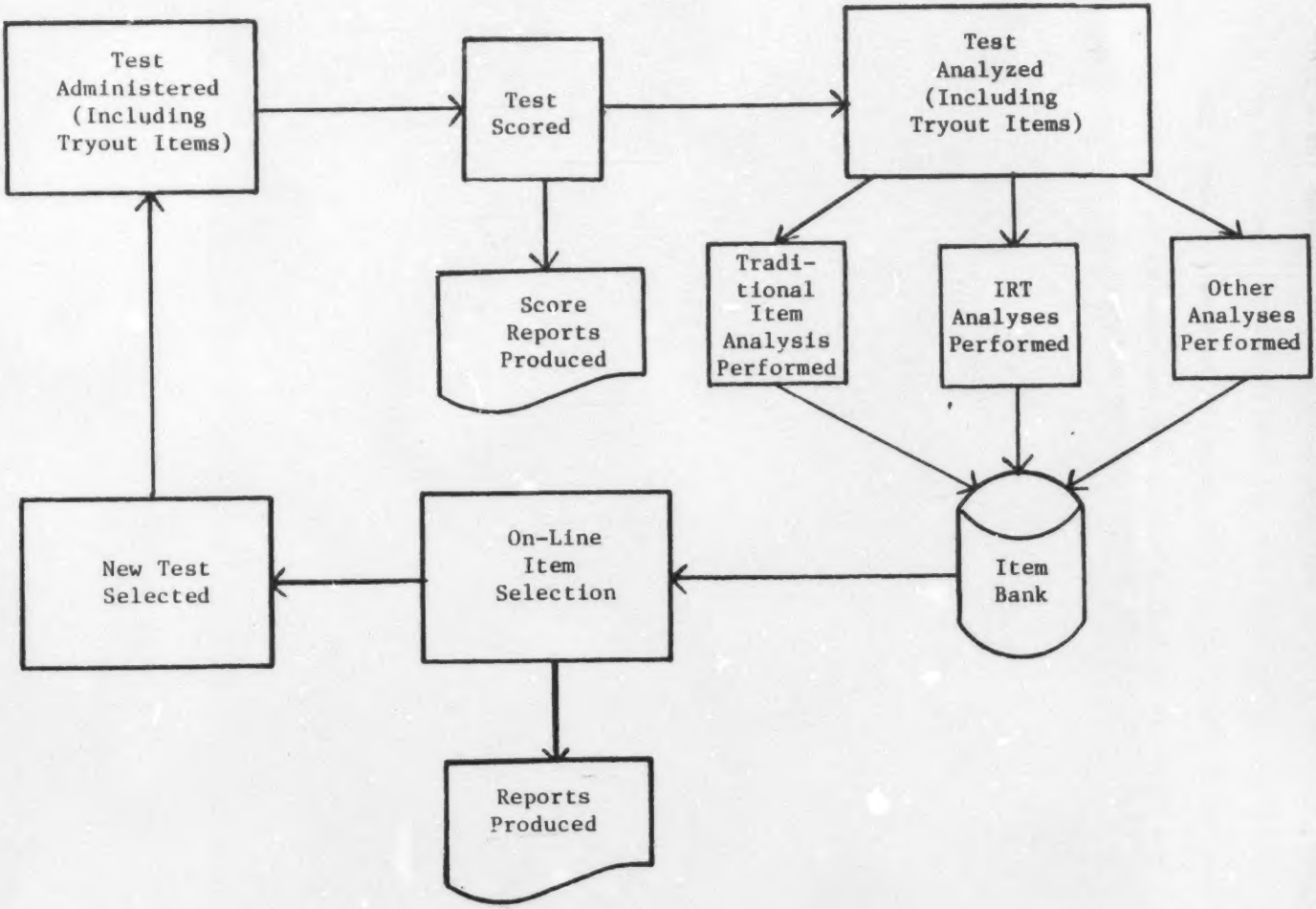

Figure 1. An overview of the CSAR System

System Design

Although the CSAR system was developed for a specific application, it was designed with an eye toward future applications as well. Wherever possible, flexibility was built into the system both in terms of what information could be stored in an item bank and in terms of how the information could be entered into the system, retrieved, displayed, and then summarized.

Discussion of the design of the CSAR system will focus on the two main components in the system. One of these components is an item bank. Issues such as how item information is stored in an item bank, what information is stored, and how it is placed there will therefore be addressed. The second component in the system is an interactive item selection program. Issues such as how item information is retrieved from an item bank, how it is displayed, and how it is summarized and reported will then be discussed.

1. How Information is Stored in the Bank

As noted earlier, the item bank is a disk file containing one data record for each administration of an item. A data record includes, among other things, an identification number that uniquely identifies the item and an administration date that, along with the identification number, uniquely identifies the data record within the bank.

In order to conserve space, item information is stored in variable length data records. Only those fields actually containing data are stored in the disk file. The length of a data record is therefore increased as data become available. As currently implemented, the maximum length of a data record is 1,000 bytes, with the average length being 200 bytes. The maximum number of such data records is 15,000.

The use of variable length data records has a second advantage. It simplifies the addition of item information not currently included in the data record. This allows the system to be kept abreast of current psychometric developments. It also increases the potential applications of CSAR in other areas where more, or different, information may be required.

Four item banks are currently in existence--one bank for each of the nurse licensing programs mentioned earlier. The banks vary in size. The smallest one contains item information for about 500 unique items while the largest one includes information for almost 3,500 items.

Since an item has one data record in the bank for each administration in which it appeared, the total number of data records in the bank is generally larger than the number of unique items in the bank. How much larger, of course, will depend on how often the same item may be administered. This is usually a matter of policy in licensing programs. The four nurse licensing banks contain, on the average, about 30 percent more data records than unique items.

New item information is added to a bank after each test administration. While the rate at which a bank grows will depend on the characteristics of the particular testing program, factors effecting the growth rate include the number of items in each examination, how often the examinations are administered, and how many experimental items are included in each. Again, these factors may be a matter of policy. The average number of data records added to the four nurse licensing banks each year is 1,000.

2. What Information is Stored in the Bank

Three kinds of item information are currently stored in an item bank: non-statistical information such as the position of an item in the test book and its correct answer, traditional item analysis information such as the

item's p-value and point biserial correlation, and item response theory (IRT) information such as item parameter estimates for the item and measures of fit. A summary of the specific item information currently stored in an item bank is given in Table 1. A more detailed discussion of this information can be found in Appendix A.

At the present time, information from the one-parameter logistic (Rasch) model is included in a data record. This IRT model was selected because it was the simplest of the currently available models that met the needs of the licensing programs for which CSAR was developed. Specifically, no vertical equating of examinations was required, and the difficulty level of items within examinations suggested a minimal amount of guessing behavior. However, the CSAR system can accommodate information from a two- or three-parameter IRT model with some additional programming effort.

3. How Information is Entered onto Storage

Item information is added, deleted, or updated in an item bank by two subroutines written in the PL/1 programming language. One subroutine reads, deletes, or rewrites data records in the bank. The other adds or replaces item information in an existing data record. The subroutines were designed to be inserted into existing application batch programs, since computer programs already existed to compute most of the statistical information stored in the bank. This approach eliminated the necessity of duplicating software specifically for use with the CSAR system.

The process of adding data records to an existing bank involves several steps. Initially, one "item shell" is created in the bank for each item to be added. This "skeleton" data record must contain at least an identification number and an administration date for each item, and it may also contain some or all of the item information outlined in Table 1. Item shells are usually

Table 1

Item Information Currently Stored in an Item Bank[a]

I. Non-Statistical Information
   1. Identification Number
   2. Administration Date
   3. Classification Code
   4. Situation
   5. Situation Modifier
   6. Item Type
   7. Anchor (previous administration)
   8. Usability Code
   9. Correct Answer
   10. Test Book Number
   11. Position in Test Book
   12. Form

II. Traditional Item Analysis Information
   1. Item Difficulty
   2. Item Difficulty Adjusted for Omits
   3. Item Discrimination
   4. Sample Size
   5. Distractor Analysis
      a) Item Difficulty
      b) Item Discrimination
      c) Fifth's Analyses

III. Item Response Theory Information
   1. Item Difficulty Estimate
   2. Standard Error of the Item Difficulty Estimate
   3. Standardized Mean Square Fit
   4. Chi Square Measure of Fit
   5. Discrimination Index
   6. Empirical Guessing Estimate

[a]A detailed description of each entry in Table 1 can be found in Appendix A.

created prior to the actual administration of the items to be added. The remaining information in the data record is inserted after the items have been administered and scored as various traditional and IRT analyses are completed.

When a bank is in the early stages of development, all of the non-statistical information in the item record must be key-entered into a temporary disk file and then loaded into the bank. However, as the number of items in the bank increases, much of this information will already be stored in the bank, but under different administration dates. Rather than key-entering it a second time, a system was developed that allowed the information that does not change from administration to administration to be read directly from the bank into a temporary disk file. Only non-statistical information that changes, such as the current administration date for the item and its position in the new examination, must then be key-entered. This added feature has dramatically reduced the amount of time required to add new items to the bank as well as decreasing the number of key-entry errors. At the present time, only the non-statistical information for experimental items must be key entered in its entirety.

A second method of adding and updating information in an item bank is available through the interactive item selection program. One of the screens in this program, the UPDATE screen, allows the test editor to display any data record in the bank by specifying its indentification number and administration date. If the test editor is properly authorized, information in that record can then be updated or a new data record created. The three parts of the UPDATE screen can be seen in Figures 9, 10, and 11 which appear later in this paper.

The ability to interact directly with an item bank through the UPDATE screen is an extremely useful, and potentially dangerous, feature of the CSAR system. Currently, only one individual has the authorization to make changes in existing data records or to create new ones "on-line." It should be noted that the ability to delete data records was not, however, included in the UPDATE screen. This serves as a safeguard against inadvertently detroying part or all of a bank. Data records can be deleted from a bank only in batch mode.

3. How information is accessed in the bank.

The second component in the CSAR system is an interactive item selection program. The purpose of this program is to facilitate the test construction process by allowing a test editor to access an existing item bank, to select items from it in a variety of different ways, and then to summarize information about the items selected.

The item selection program consists of 12 interactive screens identified by short names, which suggest their functions. Movement from one screen to the next is accomplished by entering the name of the next screen in the upper left-hand corner of the screen currently being displayed. Where logical patterns of movement between screens exist, a default screen name is automatically provided. The 12 screens in the interactive item selection program and the function of each are presented in Table 2 in alphabetic order by screen name.

The first screen presented to the test editor by the interactive item selection program is the SELDB screen. It allows the test editor to select the item bank (data base) with which he or she will work. A sample of the SELDB screen can be seen in Figure 2.

Once an item bank has been specified, items may be selected from that item bank in several ways. This function is performed by two screens: SELID and

Table 2

The Names and Functions of the Twelve Screens in the Interactive Item Selection Program in the CSAR System[a]

SCREEN NAME	SCREEN FUNCTION
CLUSTER	Display, Delete, and Combine Item Clusters
DISPLAY	Display and Delete Items in an Item Cluster
HELP	Display Screen Names and Their Functions
OPTION1	Specify Display Pattern and Name Clusters
OPTION2	Specify Clusters to be Summarized
SECURVE	Display Standard Error Curve for Clusters
SELDB	Select Item Bank (Database)
SELID	Select Items by Identification Number
SELITMS	Select Items by Classification Code and/or Statistics
SITSUM	Summarize Situation Information
SUMMARY	Summarize Item Clusters
UPDATE	Display, Add, or Update Items

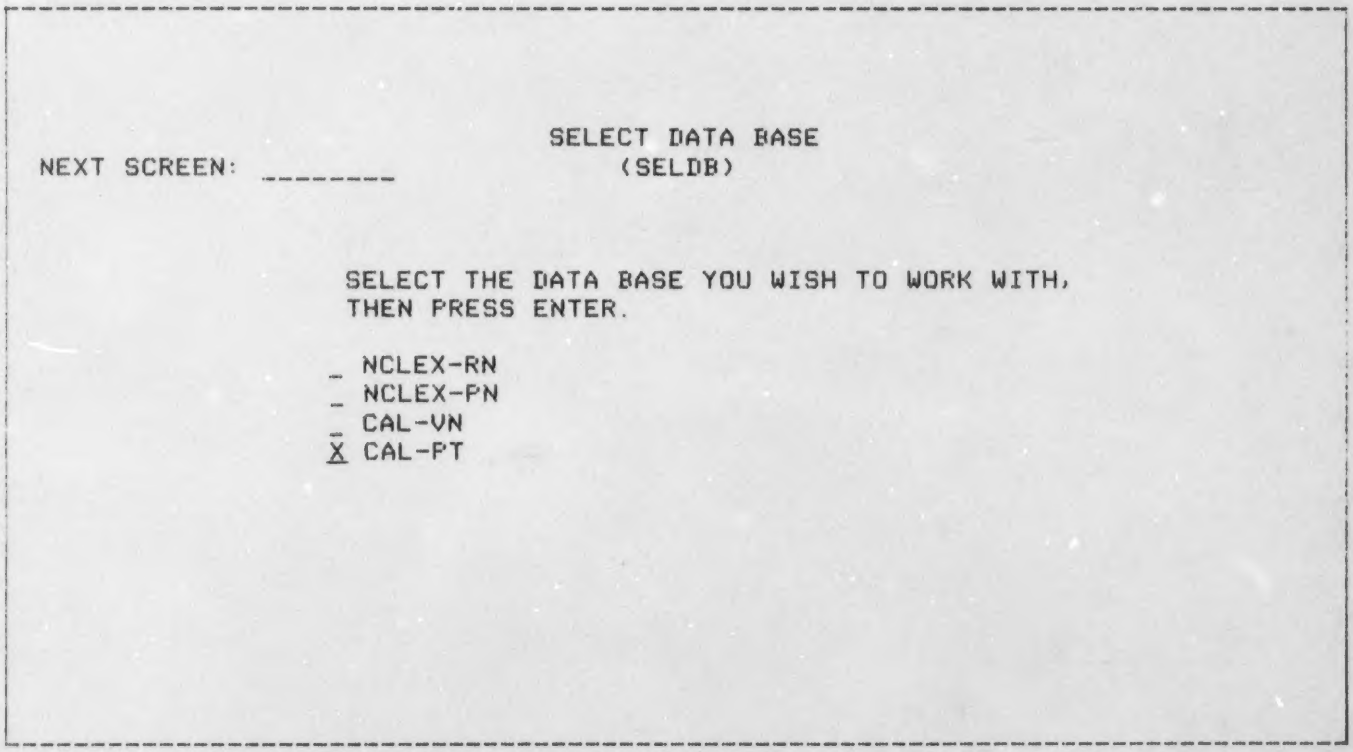

Figure 2. The SELDB screen allows the test editor to select the particular item bank (data base) with which he or she will work.

SELITMS. The SELID screen allows a test editor to select items by specific identification number or by range of numbers. The selection can be restricted to a single test administration date, or it can include all administrations of the items. A sample of the SELID screen can be found in Figure 3.

The SELITMS screen allows the test editor to select items by classification code and/or by a value or range of values for most of the entries in Table 1. The SELITMS screen has two parts, which can be seen in Figures 4 and 5. Items are selected from the bank after both parts of the screen have been completed, so that an infinitely large number of potential selection criteria exist. For example, a test editor could select from the bank all experimental items with a particular classification code and administration date. Further, the test editor could specify that only those items designated as "usable" be selected and only if certain statistical criteria were met such as a point biserial correlation coefficient greater than .30 and a b-value between ± 2.0.

After a test editor has specified selection criteria on either the SELID or SELITMS screens, the item selection program reviews all data records in the bank and stores those that meet the criteria in an "item cluster." An item cluster is any collection of data records stored under a common name. As used here, item clusters are the building blocks of examinations. They usually represent some part of the test content outline, so that several item clusters would contain all the items in one examination. Once a cluster has been created, items may be added to it or deleted from it at any time, and one or more clusters may be combined to form a new cluster. Any cluster created by a test editor remains stored in his or her library until it is deleted. Existing clusters may be viewed on the CLUSTERS screen. A sample of the CLUSTERS screen is presented in Figure 6.

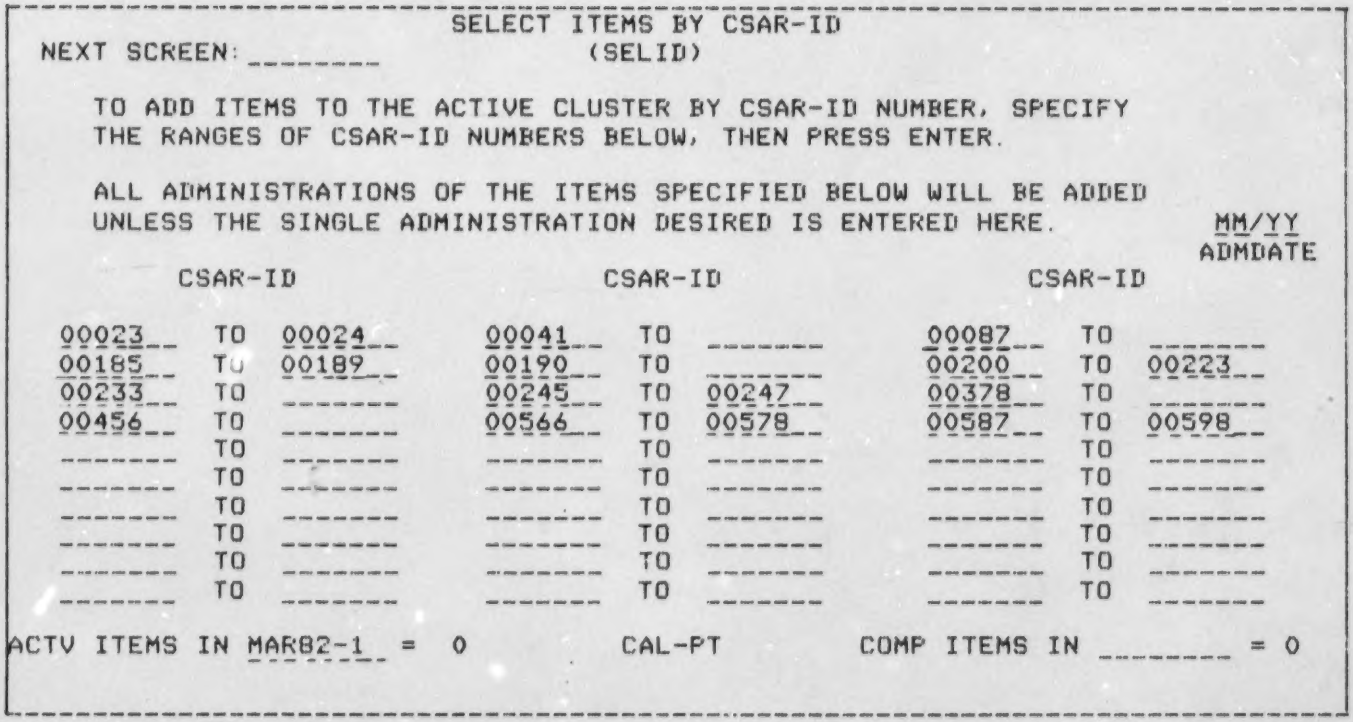
SELECT ITEMS BY CSAR-ID
NEXT SCREEN: _____ (SELID)
TO ADD ITEMS TO THE ACTIVE CLUSTER BY CSAR-ID NUMBER, SPECIFY THE RANGES OF CSAR-ID NUMBERS BELOW, THEN PRESS ENTER.
ALL ADMINISTRATIONS OF THE ITEMS SPECIFIED BELOW WILL BE ADDED UNLESS THE SINGLE ADMINISTRATION DESIRED IS ENTERED HERE.
MM/YY ADMDATE
CSAR-ID
CSAR-ID
CSAR-ID
00023__ TO 00024__
00185__ TO 00189__
00233__ TO _____
00456__ TO _____
00041__ TO _____
00190__ TO _____
00245__ TO 00247__
00566__ TO 00578__
00087__ TO _____
00200__ TO 00223__
00378__ TO _____
00587__ TO 00598__
ACTV ITEMS IN MAR82-1 = 0
CAL-PT
COMP ITEMS IN _____ = 0

Figure 3. The SELID screen allows the test editor to select items from an item bank by specific identification number or range of numbers.

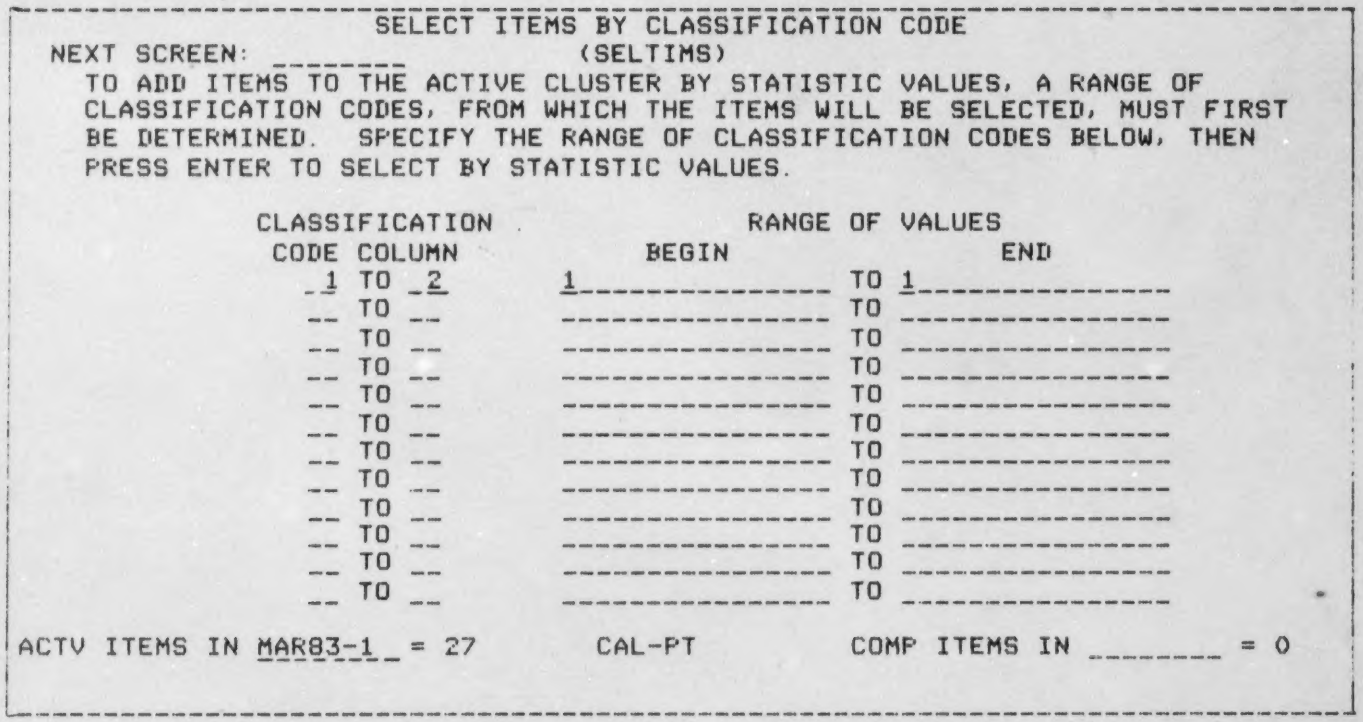
SELECT ITEMS BY CLASSIFICATION CODE
NEXT SCREEN: _____ (SELTIMS)
TO ADD ITEMS TO THE ACTIVE CLUSTER BY STATISTIC VALUES, A RANGE OF CLASSIFICATION CODES, FROM WHICH THE ITEMS WILL BE SELECTED, MUST FIRST BE DETERMINED. SPECIFY THE RANGE OF CLASSIFICATION CODES BELOW, THEN PRESS ENTER TO SELECT BY STATISTIC VALUES.
CLASSIFICATION CODE COLUMN
RANGE OF VALUES
BEGIN
END
1 TO 2
1 TO 1
ACTV ITEMS IN MAR83-1 = 27
CAL-PT
COMP ITEMS IN _____ = 0

Figure 4. Part I of the SELITMS screen allows the test editor to select items from an item bank by a specific classification code or range of codes.

```
                    SELECT ITEMS BY STATISTICS
  NEXT SCREEN: ________          (SELITMS)
    SPECIFY THE BEGINNING AND ENDING VALUES FOR ANY OF THE STATISTICS
    BELOW, THEN PRESS ENTER.  ITEMS HAVING VALUES WITHIN THE RANGES
    SPECIFIED WILL BE ADDED TO THE ACTIVE CLUSTER.

STATISTIC   NOT   BEGIN VALUE   END VALUE   STATISTIC   NOT   BEGIN VALUE   END VALUE
   ADMDATE   _      07/81         11/82       P-VALUE   _     X.XXXX        X.XXXX
   SIT       _      XXXX          XXXX        PADJ      _     X.XXXX        X.XXXX
   SITMOD    _      _             _           PBIS      _     +0.300        +1.000
   TYPE      _      I             _           N         _     XXXXX         XXXXX
   ANCHOR    _      MM/YY         MM/YY       B-VALUE   _     -2.0000       +2.0000
   USABLE    _      OK            --          SE        _     XX.XXXX       XX.XXXX
   DATA-ADQ  _      --            --          TXFIT     _     SXX.XXXX      SXX.XXXX
   ANSWER    _      _             _           WFIT      _     XXX.XXXX      XXX.XXXX
   POSITION  _      XXX           XXX         RDIS      _     X.XXXX        X.XXXX
   BOOK      _      _             _           GUESSING  _     X.XXXX        X.XXXX
   FORM      _      _             _

BEGIN CLASSIF CODE: 1                        END CLASSIF CODE: 1
ACTV ITEMS IN MAR83-1_ = 27          CAL-PT      COMP ITEMS IN ________ = 0
```

Figure 5. Part II of the SELITMS screen allows the test editor to select items from an item bank by a specific value or range of values for any of the entries above.

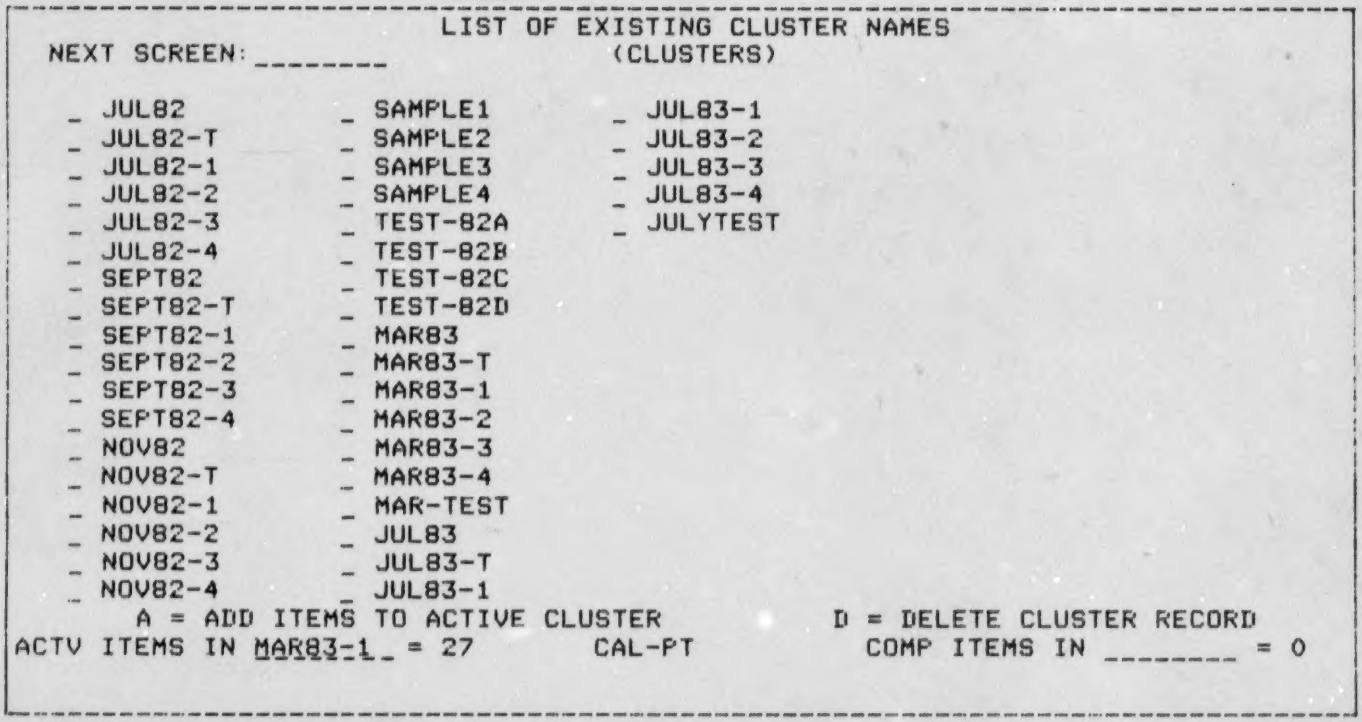

Figure 6. The CLUSTERS screen displays all cluster names currently in use.

The item cluster with which a test editor is currently working is referred to as the "active" cluster. It is specified on the OPTION1 screen, and items selected with the SELID or SELITMS screens are automatically added to the active cluster. A "comparison" cluster may also be specified. Any existing cluster may be designated as the comparison cluster. Usually, however, a comparison cluster is the counterpart of the cluster currently being built. It contains items from the same part of the test content outline but for an earlier administration of the examination. Once a comparison cluster has been identified, the Rasch item difficulty estimates for the items in this cluster can be displayed along with information about the items in the active cluster. For convenience, the number of items in both the active and comparison clusters are reported at the bottom of most screens in the program. A sample of the OPTION1 screen can be found in Figure 7.

4. How Information is Displayed

Information about items is displayed on the DISPLAY screen. Exactly what information will be presented there and the configuration in which it will appear is determined by the test editor. Selections are made from a menu on the OPTION1 screen where the cluster to be displayed, the active cluster, is also specified (see Figure 7). The information selected from the menu then appears on the DISPLAY screen in the order specified by the test editor. The items to be displayed may be sorted in either ascending or descending order, and particularly useful display patterns can be stored under a "pattern" name also specified by the test editor. The ability to store display patterns eliminates the need to re-enter this information for frequently used configurations. A sample of the DISPLAY screen appears in Figure 8.

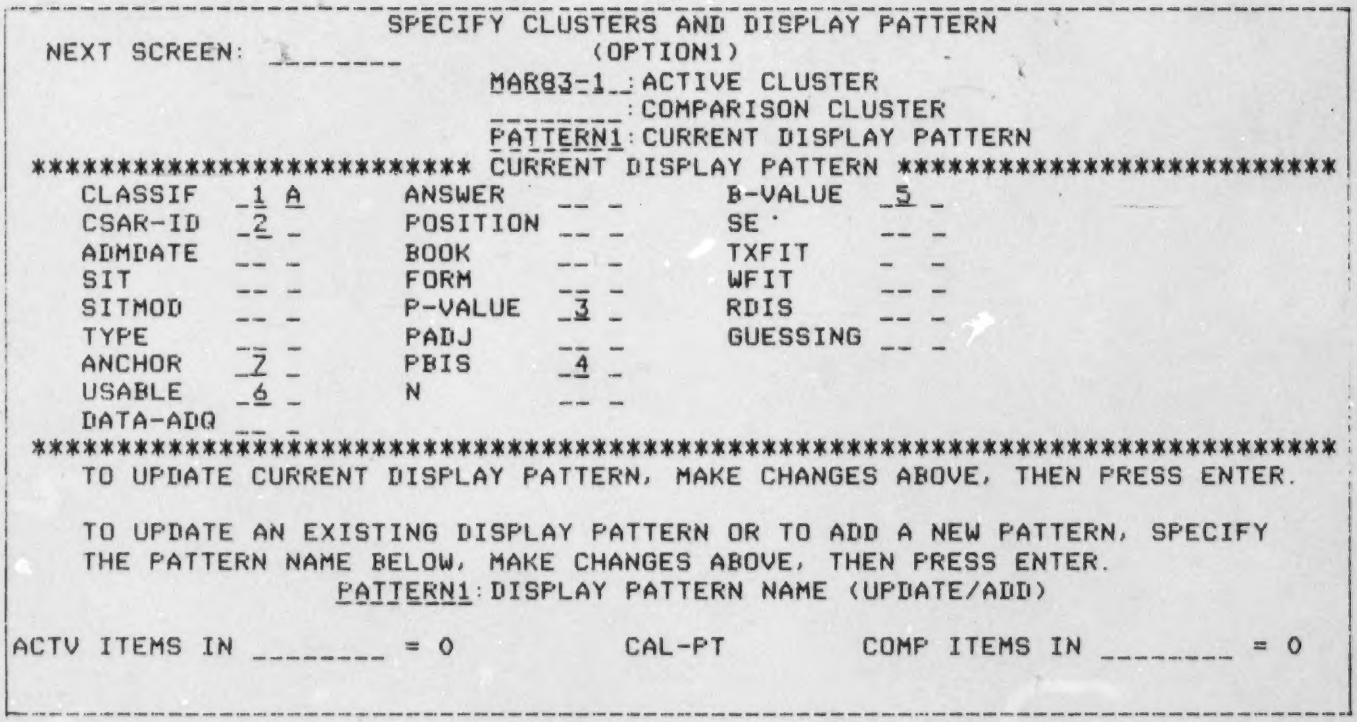

Figure 7. The OPTION1 screen allows the test editor to specify an active and comparison cluster and to specify an existing display pattern (or to create a new one).

```
NEXT SCREEN: ________        DISPLAY OF ITEM INFORMATION          NEXT ITEM 0017
 REL                                (DISPLAY)                  NEXT B-VALUE 0001
 NUM  |     CLASSIF   CSAR-ID   P-VALUE   PBIS   B-VALUE   USABLE   ANCHOR |
_ 0001 |   1A 001      00039      0.49     0.25    0.14      OK       07/82 |
_ 0002 |   1A 007      00071      0.89     0.30   -2.02      OK       09/82 |
_ 0003 |   1A 091      00076      0.87     0.35   -1.68      OK       07/82 |
_ 0004 |   1A 020      00415      0.94     0.28   -2.80      OK             |
_ 0005 |   1B 023      00016      0.79     0.29   -1.13      OK       11/82 |
_ 0006 |   1B 023      00029      0.69     0.25   -0.66      OK             |
_ 0007 |   1B 023      00031      0.85     0.47   -1.57      NO             |
_ 0008 |   1B 023      00038      0.98     0.37   -3.77      OK       07/82 |
_ 0009 |   1B 023      00066      0.82     0.22   -1.57      OK       07/82 |
_ 0010 |   1C 024      00021      0.87     0.33   -1.90      OK       03/82 |
_ 0011 |   1C 026      00015      0.65     0.20   -0.49      OK             |
_ 0012 |   1C 029      00033      0.82     0.32   -1.63      OK       11/82 |
_ 0013 |   1C 029      00046      0.79     0.29   -1.11      OK             |
_ 0014 |   1D2043      00227      0.67     0.24   -1.97      OK       09/82 |
_ 0015 |   1D2043      00417      0.89     0.28   -1.90      NO       09/82 |
_ 0016 |   1D3036      00067      0.43     0.28    0.49      OK       11/82 |
                      D = DELETE FROM THE ACTIVE CLUSTER
ACTV ITEMS IN MAR83-1 = 27          CAL-PT          COMP ITEMS IN ________= 0
 PF1: ITEMS FORWARD    PF3: B-VALUES FOWARD     PF5: DISPLAY NEXT ITEM/B-VALUE
 PF2: ITEMS BACKWARD   PF4: B-VALUES BACKWARD   ENTER: TRANSFER TO NEXT SCREEN
```

Figure 8. The DISPLAY screen displays information about the items in the active cluster. What information will be displayed and the order in which it will appear is specified on the OPTION1 screen.

Sixteen items can be viewed on the DISPLAY screen at one time. The amount of information displayed for each item depends to a large extent on the specific information chosen for display and on whether a comparison cluster was designated. The DISPLAY screen can be scrolled or paged forward or backward. If a comparison cluster is specified, the Rasch item difficulty estimates for the items in that cluster are also presented on the DISPLAY screen. These values can be independently scrolled forward and backward.

Another function of the DISPLAY screen is to delete items from the active cluster. In the process of honing down the set of items initially selected from the bank, the test editor may move back and forth between the OPTION1 and DISPLAY screens several times, each time viewing different information about the items before deleting one or more of them. Additional items may need to be selected from the bank using the SELID or SELITMS screens before a final set is identified.

A second screen on which item information is displayed is the UPDATE screen. All the information associated with a particular administration of an item can be displayed on this three-part screen. In order to view a particular data record, the test editor must enter the correct identification number for the item and a specific administration date. Traditional and IRT item analysis information is displayed as well as a distractor analysis, including p-values and point biserial correlations by distractor and an analysis by fifths. The three parts of the UPDATE screen can be seen in Figures 9, 10, and 11.

5. How Information is Summarized

When developing equivalent forms of an examination, comparisons between the clusters currently being built and their counterparts from one or more previous examinations are useful. The interactive item selection program was

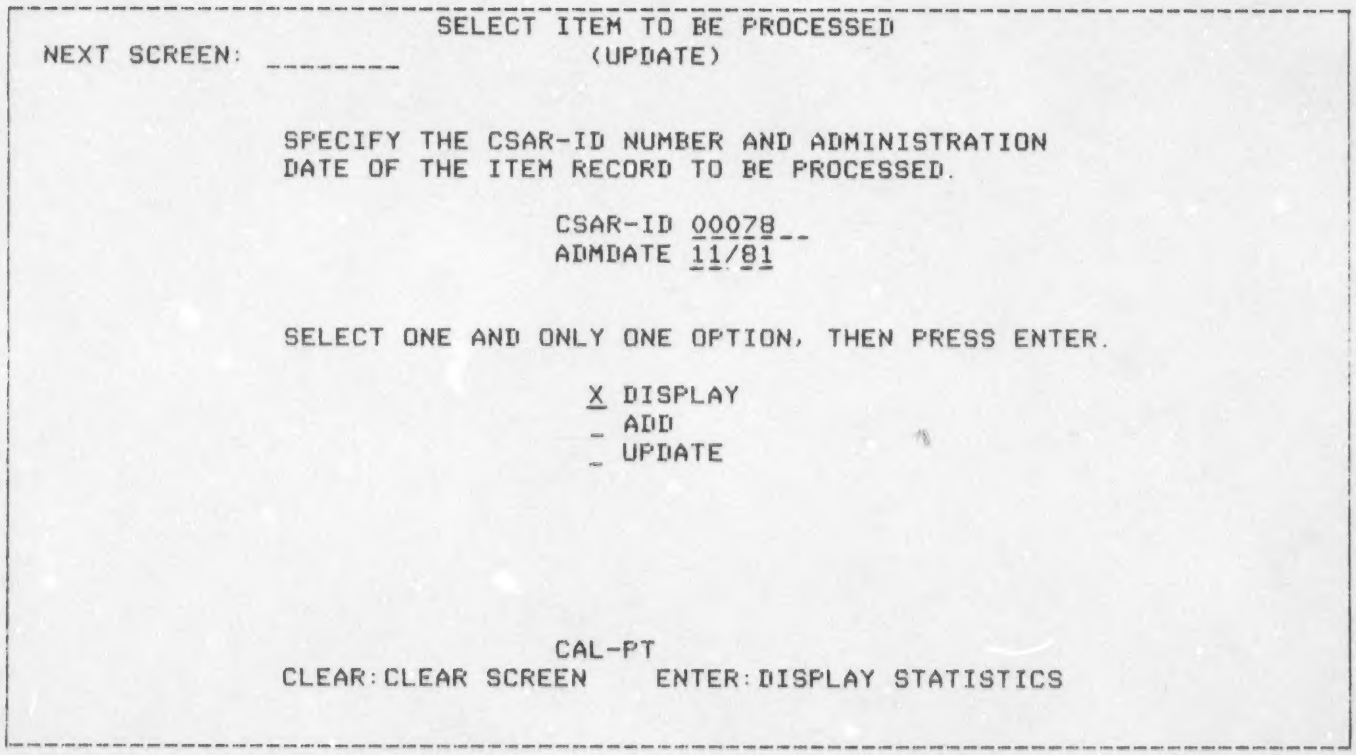

Figure 9. Part I of the UPDATE screen allows the test editor to specify the identification number of an item and a particular administration date so that the item record may then be displayed, added, or updated.

```
                          DISPLAY OF ITEM STATISTICS
NEXT SCREEN: ________           (UPDATE)

                                CSAR-ID 00073
                                ADMDATE 11/81

 CLASSIF     1D6039               ANSWER     4              B-VALUE   -2.3063
 SIT         0145                 POSITION   105            SE        00.1064
 SITMOD      $                    BOOK       2              TXFIT     +00.4136
 TYPE        R                    FORM       *              WFIT      001.4141
 ANCHOR                           P-VALUE    0.9110         RDIS      1.3999
 USABLE      OK                   PADJ       0.9113         GUESSING  0.4094
 DATA-ADQ    OK                   PBIS       +0.1954
                                  N          02820

                                CAL-PT
CLEAR:CLEAR SCREEN ENTER:NEXT SCREEN PF1:DISPLAY DISTRACTORS PF2:SELECT ITEM
```

Figure 10. Part II of the UPDATE screen displays all the item information currently stored in the bank for the particular administration of the item specified in Part I of the screen.

```
                        DISPLAY OF DISTRACTOR ANALYSIS
 NEXT SCREEN: _______             (UPDATE)

                                  CSAR-ID 00073
                                  ADMDATE 11/81
          1        2        3        4        5        6        7        8     OMIT
NSEL    00020    00174    00056    02569    00000    00000    00000             00000
 SEL       1%       6%       2%      91%       0%       0%       0%               0%

NSELVH  00000    00013    00003    00556    00000    00000    00000             00000
 SELVH     0%       2%       1%      97%       0%       0%       0%               0%
NSELH   00001    00018    00008    00526    00000    00000    00000             00000
 SELH      0%       3%       1%      95%       0%       0%       0%               0%
NSELM   00001    00037    00011    00546    00000    00000    00000             00000
 SELM      0%       6%       2%      92%       0%       0%       0%               0%
NSELL   00004    00030    00010    00485    00000    00000    00000             00000
 SELL      1%       7%       2%      90%       0%       0%       0%               0%
NSELVL  00014    00067    00024    00456    00000    00000    00000             00000
 SELVL     2%      12%       4%      81%       0%       0%       0%               0%

PBISD  -0.1673  -0.1803  -0.1282  +0.1954  +0.0000  +0.0000  +0.0000           +0.0000

                                  CAL-PT
 CLEAR:CLEAR SCREEN ENTER:NEXT SCREEN  PF1:SELECT ITEM  PF2:DISPLAY STATISTICS
```

Figure 11. Part III of the UPDATE screen displays a distractor analysis (by fifths) for the particular administration of the item specified in Part I of the screen.

designed to facilitate these kinds of comparisons. Four screens were developed to provide this kind of summary information. OPTION2, SUMMARY, SECURVE, and SITSUM. The OPTION2 screen is used to designate one or more clusters as the "new" cluster, which will then be compared to one or more clusters designated as the "old" cluster. Various comparisons between the old and new cluster are performed and reported on the three remaining screens. A sample of the OPTION2 screen can be seen in Figure 12.

The SUMMARY screen summarizes various statistical information about the old and new cluster identified on the OPTION2 screen. The SECURVE screen displays a standard error curve for each of these clusters. Values of the curve are computed directly from the Rasch model for selected ability levels, and the two curves are superimposed to facilitate comparisons. The SITSUM screen, still under development, summarizes information about the "patient situations" included in each cluster. A patient situation is much like a reading passage in a reading comprehension test. It presents information about a particular situation and is then followed by one or more questions related to the situation. Samples of the SUMMARY and SECURVE screens can be found in Figures 13 and 14, respectively.

If at anytime during an item selection session the test editor should forget the name of a screen or incorrectly spell a screen name, the item selection program automatically displays the HELP screen. The HELP screen lists the names and functions of all screens in the program. A reproduction of this screen can be seen in Figure 15. To end a session, the key word "EXIT" is entered in the field for the next screen. This key word automatically terminates the program and signs the test editor off the computer system.

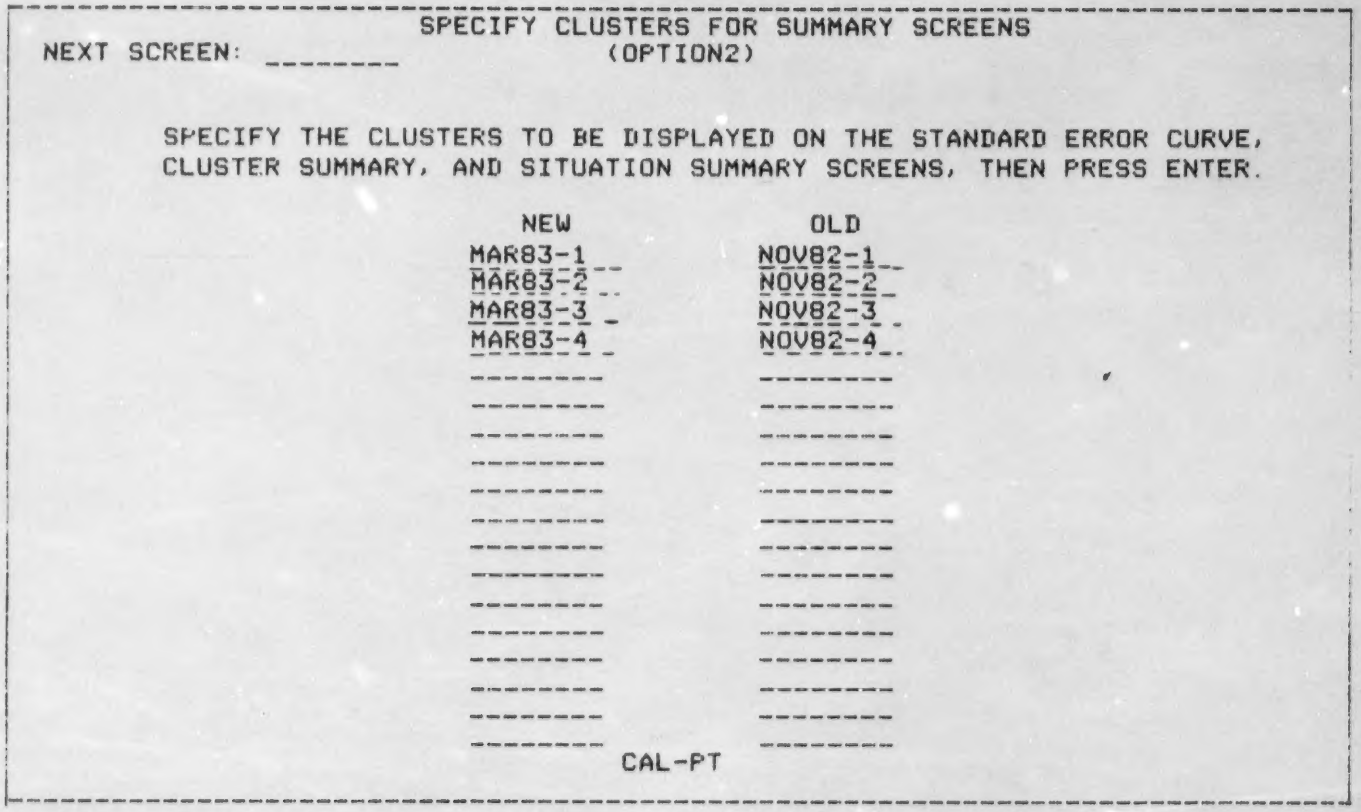

Figure 12. The OPTION2 screen allows the test editor to designate one or more existing clusters as the "new" cluster to be compared against one or more existing clusters designated as the "old" cluster. Comparisons are then displayed on the SUMMARY and SECURVE screens.

SUMMARIZE BY CLUSTER

NEXT SCREEN: _____ (SUMMARY)

	NEW	OLD	DIFF
NUMBER OF ITEMS	102	87	15
NUMBER OF ITEMS IN COMMON	57		
MEAN P-VALUE	0.72	0.67	0.05
MEAN PADJ	0.71	0.66	0.05
MEAN POINT BISERIAL	0.23	0.27	-.04
MEAN B-VALUE	-1.02	-1.02	0.00
MEAN STANDARD ERROR OF B	0.16	0.14	0.02
MEAN TXFIT	0.38	0.54	-0.16
MEAN WFIT	15.43	16.76	-1.33
MEAN RASCH DISCRIMINATION	0.99	1.01	-0.02
MEAN EMPIRICAL GUESSING	0.37	0.35	0.02

	1	2	3	4	5	1	2	3	4	5
CORRECT ANSWERS	22	32	24	24	0	17	19	26	25	0

ITEM FIT INDEX	GOOD		FAIR		POOR	GOOD		FAIR		POOR
	12		49		41	24		43		20
		OK		NO		OK			NO	
DATA ADEQUACY		83		19		86			1	
ITEM USABILITY		102		0		86			1	

CAL-PT

Figure 13. The SUMMARY screen allows the test editor to make various comparisons between the old and new clusters specified on the OPTION2 screen.

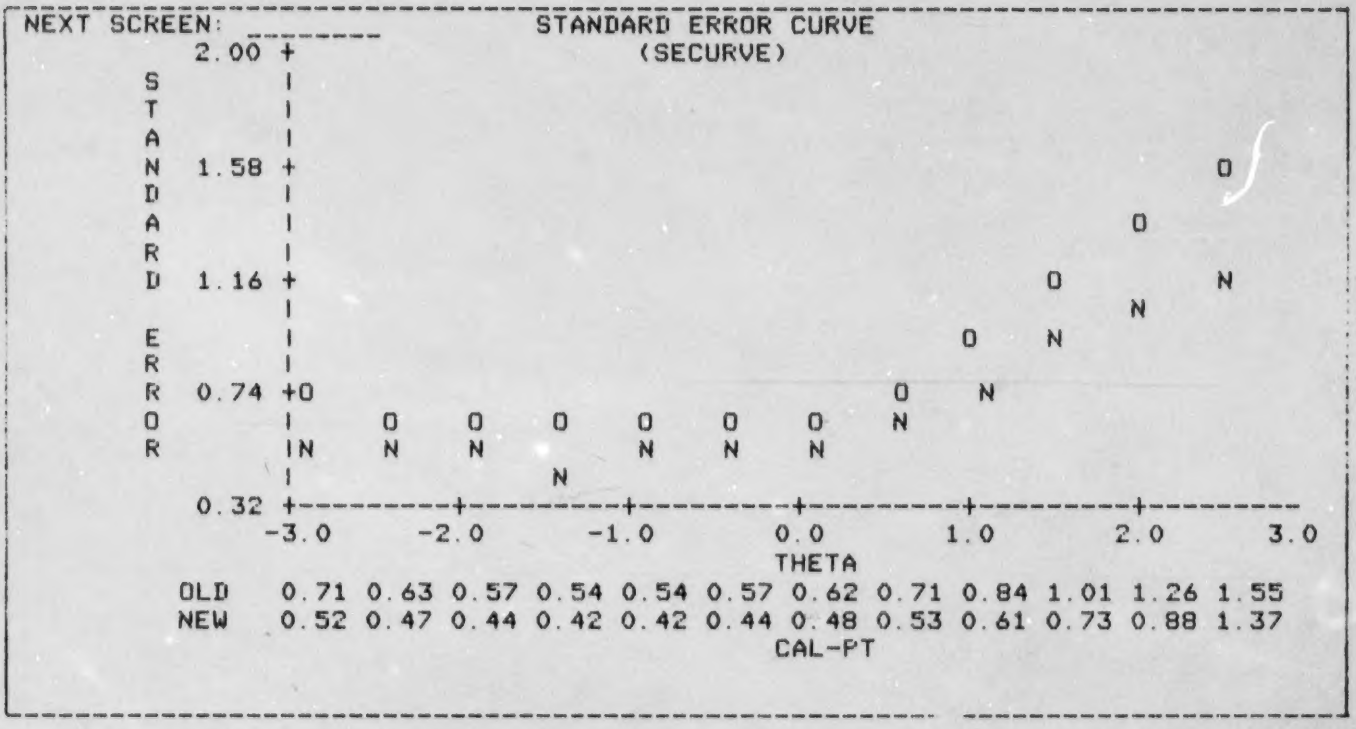

Figure 14. The SECURVE screen displays a standard error curve based on the Rasch model for the old and new clusters specified on the OPTION2 screen.

```
                    NAMES AND FUNCTIONS OF SCREENS
NEXT SCREEN: _______                (HELP)
        SELECT THE SCREEN YOU WISH TO SEE NEXT, THEN PRESS ENTER.


CLUSTERS    (DISPLAY AND DELETE CLUSTER NAMES)
DISPLAY     (DISPLAY AND DELETE ITEMS IN ACTIVE CLUSTER)
OPTION1     (SPECIFY DISPLAY PATTERN, ACTIVE AND COMPARISON CLUSTERS)
OPTION2     (SPECIFY OLD AND NEW CLUSTERS FOR SUMMARY SCREENS)
SECURVE     (DISPLAY STANDARD ERROR CURVE FOR OLD AND NEW CLUSTERS)
SELDB       (SELECT DATA BASE)
SELID       (SELECT ITEMS BY CSAR-ID)
SELITMS     (SELECT ITEMS BY CLASSIFICATION CODE AND STATISTICS)
SITSUM      (DISPLAY SITUATION INFORMATION IN OLD AND NEW CLUSTERS)
SUMMARY     (DISPLAY DIFFERENCES BETWEEN OLD AND NEW CLUSTERS)
UPDATE      (DISPLAY/ADD/UPDATE ITEM RECORD)


HELP        (DISPLAY NAMES AND FUNCTIONS OF SCREENS)
EXIT        (END TRANSACTION)
```

Figure 15. The HELP screen displays the names and functions of all screens in the interactive item selection program.

As the system is currently implemented, a printed copy of any of the screens in the item selection program can be produced by a special function key, which triggers a small on-line printer. This capability has speed as an advantage, since printed copies of any information displayed by the item selection program are available immediately. The disadvantage of this method of reporting is that it can become tedious if lengthy reports must be produced. More extensive reporting capabilities are currently being investigated.

Recommendations

In the last section of this paper, some recommendations will be offered to others who might be interested in developing an item banking system similar to CSAR. The discussion will focus mainly on design issues related to the two components of the system. Some desirable features to include in such a system will also be described. First, some general comments on system design will be presented.

The most important part of the process of developing an item banking system is designing it. If such a venture is being considered, a lot of time should be spent on this activity. The CSAR system was particularly difficult to design because of the need to address the specific requirements of several nurse licensing examination programs and to meet these needs in ways that maximized the potential for future applications of the system. Such an approach was necessary in order to make the system cost-effective, since development and implementation costs were in the neighborhood of $100,000.

One other comment on system design seems in order. Adjustments to the system will be necessary after implementation, regardless of how carefully the system was designed.. When the system goes into a production mode, unforseen problems will arise. A period of fine-tuning is therefore inevitable during the initial implementation of the system. If at all possible, the development of the system should be coordinated so that at least one item bank is available for system checking when the item selection program is completed. This will greatly facilitate the location of any problems or inconsistencies in this program, especially if the test is performed by someone other than the programmer who wrote it.

The Item Bank

The first component of the CSAR system is one or more item banks. As the system is currently implemented, one data record is stored in an item bank for each administration of an item. Each record includes, among other things, a unique identification number for the item and an administration date. This approach has several advantages, including simplicity. However, it has several disadvantages, too. Item information common to all data records for the same item must be duplicated on every record for that item. This includes information such as the correct answer for the item and its classification code. Unless great care is taken, discrepancies between records can be introduced, which require time and effort to resolve. This is particularly troublesome when the classification scheme for items is general enough that an item might fall into more than one category. There is a temptation to change the item's classification code to meet the requirements of the test plan for a particular administration. However, such changes are discouraged because they require updating the data records for all previous administrations of the item in order to maintain consistency.

The problem of duplicate information and effort might be avoided by having only one data record in the bank for each item. The information that does not change from administration to administration would then be entered only once on this record. Although such an approach has other problems associated with it, it would be worth exploring further before finalizing the design of a new system.

Along the same lines, some care must be exercised in selecting a scheme for classifying items in the bank. The process of reviewing items and assigning code values to them is very time-consuming and, as mentioned above, error-prone. Once the item bank contains enough items to select examinations exclusively from it, the process of entering classification information should definitely be automated wherever possible. In addition, the option of having more than one classification code for an item should be considered if a primary and a secondary classification of items would be useful in test construction.

A problem frequently encountered when selecting items from an item bank is that once a set has been identified, one item in it may give away the answer to another. This problem of "contamination" is serious enough to recommend that some method of handling it be included in the design of a new system. However, no simple solutions come to mind, although having item text available for review while making the selections might alleviate the problem somewhat.

The Interactive Item Selection Program

The second component of the CSAR system is an interactive item selection program. The purpose of the program is to facilitate the test construction process. In order to achieve this end, the program must be "user-friendly" since it will be used primarily by test editors. This can be accomplished by keeping the functions of the screens as simple and self-explanatory as possible and by avoiding the use of computer jargon. Where possible, maintain

a consistent set of instructions throughout the program so that the same action on the part of the test editor produces the same result regardless of the screen. Where information must be supplied by the test editor, have the program supply the most frequently chosen response automatically so that the test editor need enter information only occasionally. Last, provide good program documentation, including guidelines for item selection based on the specific item information available in the bank.

A major consideration in designing an interactive program such as the one in the CSAR system is speed. People lose interest quickly if they are forced to sit inactive because of slow response time. To a certain degree, response time will be a function of the computer environment in which the program operates. However, it can be lowered by keeping the number of activities performed "on-line" to a minimum. Any operations that can be performed in batch mode should be performed there.

Other desirable features of an interactive item selection program include the ability to select items from an item bank in a variety of different ways, to display and summarize whatever information about them might be of interest, and to produce hard copy reports quickly. Additional features not included in the item selection program in the CSAR system are the ability to sort items in more than one way, more extensive report capabilities, the ability to produce a conversion table based on the Rasch model indicating the ability estimates associated with each raw score in an item set, and the capability of displaying item text and item characteristics simultaneously.

REFERENCES

Allen, M.J., & Yen, W.M. Introduction to measurement. Monterey, CA: Brooks/Cole Publishing Company, 1979.

Curry, A.R. Invariance of the Rasch model ability estimates over different collections of items. Unpublished doctoral dissertation, University of Georgia, 1977.

Horst, A.P. The difficulty of a multiple-choice test item. Journal of Educational Psychology, 1933, 24, 229-232.

Lord, F.M. A theory of test scores. Psychometric Monograph, 1952, No. 7.

Lord, F.M. The relation of test score to the trait underlying the test. Educational and Psychological Measurement, 1953, 13, 517-548.

Lord, F.M. Applications of item response theory to practical testing problems. Hillsdale, NJ: Lawrence Erlbaum Associates, 1980.

Lord, F.M., & Novick, M.R. Statistical theories of mental test scores. Reading, MA: Addison-Wesley Publishing Company, 1968.

Rasch, G. Probabilistic models for some intelligence and attainment tests. Copenhagen, Denmark: Danmarks Paedogogiske Institut, 1960.

Rasch, G. An individual approach to item analysis. In P.F. Lazarsfeld and N.W. Henry (Eds.), Readings in Mathematical Social Science. Chicago, IL: Science Research Associates, 1966, 89-908.

Rentz, R.R., & Bashaw, W.L. Equating reading tests with the Rasch model, Volume I final report, Volume II technical reference tables. Athens, GA: University of Georgia, Educational Research Laboratory, 1975.

Warm, T.A. A primer of item response theory (CG-941278). U.S. Coast Guard Institute: P.O. Substation 18, Oklahoma City, Oklahoma, December 1978.

Wood, R.L., Wingersky, M.S., & Lord, F.M. A computer program for estimating examinee ability and item characteristic curve parameters (ETS RM 76-6). Princeton, NJ: Educational Testing Service, 1976.

Wright, B.D. Sample-free test calibration and person measurement. Proceedings of the 1967 Invitational Conference on Testing Problems. Princeton, NJ: Educational Testing Service, 1968, 85-101.

Wright, B.D. Solving measurement problems with the Rasch model. Journal of Educational Measurement, 1977, 14, 97-116.

Wright, B.D., & Douglas, G.A. Best procedures for sample-free item analysis. Applied Psychological Measurement, 1977, 1, 281-295.

Wright, B.D., Mead, R.J., & Bell, S.R. BICAL: Calibrating items with the Rasch model (Research Memorandum 23B). Chicago, IL: Statistical Laboratory, Department of Education, University of Chicago, 1979.

Wright, B.D., & Stone, M.H. Best test design: Rasch Measurement. Chicago, IL: Mesa Press, 1979.

Yen, W.M. Using simulation results to choose a latent trait model. Applied Psychological Measurement, 1981, 5, 245-262.

APPENDIX A

Item Information Currently Stored in the Item Bank

Three kinds of information are currently stored in each item bank: non-statistical information, traditional item analysis information, and item response theory information. A description of the specific information within each category follows next. The symbols used to reference information within the CSAR system are given in parenthesis.

I. Non-Statistical Item Information (CSAR-ID)

1. Identification Number (CSAR-ID)

   A unique identification number is assigned to each test item in a given item bank. The number may be up to seven alphanumeric characters in length. Each time the item is used, it is assigned the same identification number, allowing it to be tracked across test administrations.

2. Administration Date (ADMDATE)

   A four-digit numeric code of the form MMYY is used to identify the administration in which an item appeared.

3. Classification Code (CLASSIF)

   The classification code is used to reference an item to a test plan, content outline, or set of objectives. It may be up to 16 digits in length, containing alphanumeric characters. The classification code can be used as an alternate index for retrieving items from the bank.

4. Situation (SIT)

   A three-digit numeric code is used to identify the patient situation with which an item is associated. A patient situation is much like a reading passage in a reading comprehension test. It presents information about a particular situation and is then followed

by one or more test questions. All items related to the same patient situation are assigned the same situation code.

5. Situation Modifier (SITMOD)

   Sometimes a patient situation is modified by providing additional information. When this occurs, the appended information is referred to as a "situation modifier." Items associated with a particular situation modifier are then assigned a one-digit situation modifier code as well as the same situation code.

6. Item Type (TYPE)

   A one-digit code is used to indicate whether an item contributed to candidate licensure scores ("R" = real) or whether it was included in an administration for experimental purposes only ("T" = tryout).

7. Anchor (ANCHOR)

   A four-digit code of the form MMYY is assigned to those items within an administration selected for equating purposes. Included among the selection criteria for anchor items is item difficulty, fit to the model, and content coverage.

8. Usability Code (USABLE)

   A two-digit code is used to identify an item that, for one reason or another, is no longer eligible for inclusion in future examinations. Reasons for ineligibility include a breach in security such as a stolen test book, official release of the item to the public, a change in the test plan, or poor item quality.

9. Correct Answer (ANSWER)

   A one-digit alphanumeric code is used to identify the correct answer option for an item.

10. Test Book Number (BOOK)

A one-digit code is used to identify the particular test book in which an item appeared.

11. Position in Test Book (POSITION)

A three-digit numeric code is used to identify the location of an item in the test book in which it appeared. This value represents the actual number printed in the test book for that item. The purpose of this code is to facilitate the location of the text of an item.

12. Form (FORM)

A one-digit alpha code is used to identify the particular experimental form in which a tryout item appeared. In the licensing programs around which CSAR was developed, experimental items may be included in one or more test books, or they may appear in a separate test book administered at the end of testing. In either case, multiple forms are generally administered.

II. Traditional Item Analysis Information

1. Item Difficulty (P-VALUE)

The ratio of the number of candidates who got the item correct and the number tested.

2. Item Difficulty Adjusted for Omits (PADJ)

The ratio of the number of candidates who got the item correct and the number who attempted it.

3. Item Discrimination (PBIS)

The point biserial correlation between the score a candidate earned on an item and his or her total test score.

4. Sample Size (N)

The number of candidates in the sample.

5. Distractor Analyses

The number and percent of candidates selecting each answer option and the point biserial correlation between each answer option and the total test score are stored in the bank as well as an analysis by fifths.

III. Item Response Theory Information

1. Item Difficulty Estimate (B-VALUE)

The item response theory model currently in use with CSAR is the one-parameter logistic (Rasch) model, given by the following expression:

$$P_i(\theta) = \frac{1}{1 + e^{-(\theta - b_i)}} .$$

Here $P_i(\theta)$ is the probability that a candidate with trait value θ will get item i correct. In this case, b_i is the only item parameter in the model, and it is referred to as the item's difficulty estimate (Rasch, 1960, 1966; Wright & Stone, 1979). An estimate of item difficulty is obtained by first scoring the item either right or wrong and then applying a maximum likelihood estimation routine such as the one in LOGIST (Wood, Wingersky & Lord, 1976).

2. Standard Error of the Item Difficulty Estimate (SE)

Associated with each item difficulty estimate is a standard error of that estimate. This statistic is computed directly from the model as follows:

$$SE(b_i) = \sum_{i=1}^{N} \left[\hat{P}_i(\hat{\theta})\ (1 - \hat{P}_i(\hat{\theta}))\right]^{-1/2}$$

3. Standardized Mean Square Fit (TXFIT)

Several measures of fit to the model are included in the bank. The first is referred to as the "Texas Fit Statistics" (TXFIT) because of its use with the Texas Assessment of Basic Skills Project (TABS) conducted by CTB in 1980-81. It is a mean square between groups statistic and is given as follows:

$$Q_{ti} = \sum_{j=1}^{J} \left[\frac{N_j (O_{ij} - E_{ij})^2}{E_{ij}\,(1-E_{ij})} \cdot \frac{L}{(J-1)(L-1)} \right] , \text{ where}$$

$$E_{ij} = \sum_{k \varepsilon j}^{Nj} \hat{P}_i(\hat{\theta}_k) .$$

Here i = 1,L items, j = 1,J groups, and N_j is the number of candidates in group j. O_{ij} is the observed proportion of candidates in group N_j who correctly answered item i. Q_{ti} is then standardized as follows:

$$TXFIT = t \cdot (Q_{ti})^{1/3} - t + 1/t , \text{ where}$$

$$t = [4.5(J - 1)]^{1/2} .$$

To calculate this statistic, candidates are rank ordered on the basis of their trait estimates and then divided into, in this instance six, cells with approximately equal numbers in each. Q_{ti} is then computed by summing across the cells and is then expressed in standardized form (Wright, Mead, and Bell, 1979).

4. Chi Square Measured Fit (WFIT)

A second statistic typically used to measure fit was also included in the bank. This fit measure represents a modification of

the following statistic studied by Yen (1981, p. 246):

$$Q_{wi} = \sum_{j=1}^{J} \frac{N_j(O_{ij} - E_{ij})^2}{E_{ij}\ (1-E_{ij})} \quad .$$

Here N_j, O_{ij}, and E_{ij} are defined as before. Since Q_{wi} is a function of sample size, it is rescaled as follows in order to allow comparisons across tests:

$$WFIT = Q_{wi}\ [1000/N].$$

Again, N is the total number of candidates used in the computations.

5. Discrimination Index (RDIS)

A third indicator of fit included in the bank is an index of item discrimination referred to here as the Rasch Discrimination Index. It is a least squares estimate of the slope of the linearly transformed observed item characteristic curve for an item (Curry, 1977; Rentz & Bashaw, 1975), and it is given by the following expressions:

$$RDIS = 1.0 - \frac{\sum_{j=1}^{J} (\bar{\hat{\theta}}_j)(E_{ij} - O_{ij})}{\sum_{j=1}^{J} (\bar{\hat{\theta}}_j)^2\ \hat{\sigma}_j^2}, \quad \text{where}$$

$$\bar{\hat{\theta}}_j = \frac{1}{N_j} \sum_{k \epsilon j}^{N_j} \hat{\theta}_k \ , \text{ and}$$

$$\hat{\sigma}_j^2 = \sum_{k \epsilon j}^{N_j} \hat{P}_1(\hat{\theta}_k)\ [1 - \hat{P}_1(\hat{\theta}_k)] \ .$$

Again, N_j, O_{ij}, and E_{ij} are defined as before, and J is six. Theoretically, the value of this index should be 1.00 for items that fit the Rasch model, although values within the range of .80 to 1.20 may be considered similar enough for reasonable fit (Curry, 1977; Rentz & Bashaw, 1975).

6. Empirical Guessing Estimate (GUESSING)

The last measure of fit currently included in CSAR is an estimate of the amount of guessing behavior elicited by the item (Horst, 1933). It is given by:

$$\hat{P}_g = \frac{p - \hat{K}}{1 - \hat{K}} \quad , \qquad \text{where}$$

$$\hat{K} = p \dot{-} a_2 \quad .$$

Here $\hat{K}$ is an estimate of the proportion of candidates who actually know the correct answer to the item, and it is computed using p, the proportion who got the item correct, and a_2, the proportion choosing the most popular (incorrect) answer option. Research investigating the validity of this statistic is currently underway.